Evolved *Escherichia coli* Strains for Amplified, Functional Expression of Membrane Proteins

Nadia Gul†, Daniel M. Linares†, Franz Y. Ho† and Bert Poolman

Department of Biochemistry, University of Groningen, Groningen Biomolecular Sciences and Biotechnology Institute, Netherlands Proteomics Centre and Zernike Institute for Advanced Materials, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Correspondence to Bert Poolman: b.poolman@rug.nl
http://dx.doi.org/10.1016/j.jmb.2013.09.009
Edited by I B. Holland

Abstract

The major barrier to the physical characterization and structure determination of membrane proteins is low protein yield and/or low functionality in recombinant expression. The enteric bacterium *Escherichia coli* is the most widely employed organism for producing recombinant proteins. Beside several advantages of this expression host, one major drawback is that the protein of interest does not always adopt its native conformation and may end up in large insoluble aggregates. We describe a robust strategy to increase the likelihood of overexpressing membrane proteins in a functional state. The method involves fusion in tandem of green fluorescent protein and the erythromycin resistance protein (23S ribosomal RNA adenine N-6 methyltransferase, ErmC) to the C-terminus of a target membrane protein. The fluorescence of green fluorescent protein is used to report the folding state of the target protein, whereas ErmC is used to select for increased expression. By gradually increasing the erythromycin concentration of the medium and testing different membrane protein targets, we obtained a number of evolved strains of which four (NG2, NG3, NG5 and NG6) were characterized and their genome was fully sequenced. Strikingly, each of the strains carried a mutation in the *hns* gene, whose product is involved in genome organization and transcriptional silencing. The degree of expression of (membrane) proteins correlates with the severity of the *hns* mutation, but cells in which *hns* was deleted showed an intermediate expression performance. We propose that (partial) removal of the transcriptional silencing mechanism changes the levels of proteins essential for the functional overexpression of membrane proteins.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

Membrane proteins play a crucial role in a wide range of physiological processes including metabolic energy conservation, nutrient uptake, detoxification, regulation of cell volume and ion homeostasis and signaling of environmental conditions. Membrane proteins represent 20–30% of the proteome of prokaryotic and eukaryotic organisms, and they comprise more than 70% of the current drug targets [1–4]. However, our understanding of function–structure relationships of membrane proteins is low as compared to water-soluble proteins [5]. This imbalance has been attributed to the unparalleled difficulties of obtaining sufficient amounts of protein for functional and structural studies. It has been noted that the most optimal host for overexpression is the native source [6,7]. However, this is not a feasible strategy for most membrane proteins since either they are not expressed at a high enough level or a suitable self-expression technology is not available [8]. We thus need new and improved heterologous expression systems. For large-scale production of proteins, bacterial and fungal expression hosts such as *Escherichia coli*, *Lactococcus lactis*, *Saccharomyces cerevisiae* and *Pichia pastoris* are often preferred [9–11]. Protein synthesis rates are generally much faster in prokaryotes than in eukaryotes [12], which can be advantageous but also cause problems in the downstream steps of membrane protein biogenesis [7]. Although every organism has its pros and cons for the production of complex (multi-domain) membrane proteins, *E. coli* remains the most widely used host for the heterologous expression of proteins in high quantities [13–15]. The genetics and the fundamental understanding of transcription, translation...
and protein folding are far better characterized than for any other microorganism.

The challenge of overproducing membrane proteins in a functionally competent state is enormous. The available literature suggest that many membrane proteins can be expressed in reasonable amounts, but often they fail to get inserted into the membrane in their native conformations and are non-functional [16,17]. The difficulties to overproduce functional membrane proteins have been attributed to factors such as (i) toxicity of the proteins to the host cell [18], (ii) mismatch in codon usage and tRNA abundance [19], (iii) limitation in the availability of precursors for protein synthesis (e.g., amino acids) [20], (iv) saturation of the membrane protein insertion machinery [21,22], (v) insufficient chaperone and foldase activity [23,24] and (vi) limited membrane space to accommodate extra protein [25]. Compared to soluble proteins, membrane proteins have a more complex biogenesis pathway and bottlenecks in their synthesis may arise from improper targeting, membrane insertion and/or folding [16,17]. The level of correctly folded membrane protein can often be enhanced by systematic optimization of several parameters such as (i) directed evolution or engineering of the expression host [26,27]; (ii) tuning of expression conditions such as temperature, growth medium and inducer strength [20]; and (iii) modification of target proteins [28–30]. Such screening is often performed by trial and error and is time consuming. More efficient and target-specific screening methods are needed to tackle the enormous expression space [31].

To date, one can simply screen for expression of membrane proteins in an overall well-folded state by fusion of green fluorescent protein (GFP) or derivative as a reporter to the C-terminus of the target protein. If the target protein properly, the C-terminal GFP is likely to fold and mature to a fluorescent protein [32,33]. In contrast, a target protein that misfolds and aggregates during protein synthesis will interfere with the folding of GFP. Hence, the GFP will not fold correctly and will not be fluorescent. Folded and misfolded membrane protein-GFP fusions also can be readily discriminated on SDS-PAGE and immunoblots. Correctly folded GFP is SDS resistant and migrates faster than SDS-solubilized misfolded GFP; the difference in migration is also observed when GFP is fused to the C-terminus of a membrane protein [34–37]. We previously reported that the in vivo activity of GFP-tagged glutamate transporter GltP from E. coli and GFP-tagged lactose transporters LacSΔIIA from Streptococcus thermophilus and LacY from E. coli correlates with the corresponding in-cell and in-gel GFP fluorescence of the proteins [36]. This suggests that, at least for these proteins, the GFP fluorescence can be used as indicator of their in vivo activity, which we here use as indicator of functional expression in E. coli.

In this study, we use an evolutionary strategy to optimize E. coli as a host for the production of functional membrane protein. We demonstrate that the combination of the validated reporter GFP and the erythromycin resistance protein (ErmC) can be used to select cells with increased expression. We obtained a number of evolved strains, whose genome was fully sequenced. Surprisingly, with different non-homologous membrane proteins from different species as target, we obtained mutations in the gene (hns) that encodes the generic transcriptional silencer H-NS. In the second part of the paper, we benchmark the expression performance of the evolved strains against hns knockout strains.

Results

Choice of antibiotic resistance marker

The target (membrane) protein was fused to a C-terminal selectable marker that confers a drug resistance phenotype, when located in the cytoplasm. In the initial studies, we screened various antibiotic resistance markers, including the chloramphenicol (CAT), the tetracycline (TetA) and the kanamycin (Aph) resistance proteins as fusion partners for selection, but none of these markers proved useful in E. coli. Either the fusion partners were degraded or they dragged the target protein into a misfolded state. Because each of these antibiotic resistance proteins form an oligomeric complex (CAT and TetA are trimeric proteins and Aph is homodimeric), we reasoned that their quaternary structure and/or cellular location (TetA is a membrane protein) posed a problem, in particular, when the membrane protein target would form an oligomer as well. We then identified the erythromycin resistance protein (ErmC) as a monomeric, soluble antibiotic resistance protein [37]. Fusions of membrane proteins with ErmC proved to be highly stable.

Choice of expression host

The first requirement for the suitability of ErmC fusions is an expression host highly sensitive to erythromycin. We tested many strains, and in general, E. coli strains are relatively insensitive to erythromycin. We then evaluated isogenic strains with different capacity to excrete hydrophobic compounds, including antibiotics, that are E. coli BW25113, E. coli BW25113A (ΔacrB) and BW25113B (ΔacrB ΔermE ΔmdIa Δkan) [38]. The differences in growth in erythromycin-containing media were marked. E. coli BW25113 could grow in up to 100 μg/ml of erythromycin, albeit that the final yield was reduced. By comparison, E. coli BW25113B was completely inhibited at 10 μg/ml erythromycin, whereas E. coli BW25113A displayed an intermediate phenotype.

Evolved Escherichia coli for protein expression

137
We selected *E. coli* BW25113B as expression host for the (membrane) protein-GFP-ErmC fusions, hereafter referred to as MP-GFP-ErmC. The selection system is based on the resistance to erythromycin obtained from ErmC. The folding state of the target membrane protein is reported by the GFP fluorescence. Therefore, a properly folded membrane protein-GFP-ErmC fusion can confer resistance to erythromycin in media for host survival (Fig. 1a).

MP-GFP-ErmC fusions were made with LacSΔIIA, BcaP, GltP, Sav1866, PacL, YidC and OxaA as target membrane protein. We observed two differently migrating protein species on immunoblots, the lower one being fluorescent in SDS-polyacrylamide gels (Fig. S2). These data are consistent with previous observations on MP-GFP fusions and show that a fraction of the target protein is well folded and that another fraction is misfolded [36,37]. Thus, GFP can be used as a reporter of the folding state of membrane proteins in dual fusions with ErmC. We also made fusions of GFP-ErmC with the water-soluble proteins GerAC (spore germination factor) and LmoSBP (substrate-binding protein), as they express well and proved valuable in screening the initial selection conditions (vide infra).

To determine the optimal evolution screening conditions, we performed several small-scale expression studies. As variables, we chose eight candidate membrane proteins, two *E. coli* hosts (MC1061 and...
BW25113), five temperatures (20, 25, 30, 37 and 42 °C) and various inducer (L-arabinose) concentrations. The expression of the fusion proteins was evaluated by whole-cell and in-gel fluorescence measurements and anti-His immunoblots. Figure S2 shows data from a representative set of experiments. For all the membrane proteins tested, the expression profiles were similar for the two E. coli strains (data not shown). The culture temperature after induction proved to be important for functional overexpression. The optimal temperature for high-level functional expression (in terms of fluorescence) was either 20 or 25 °C. While at 37 and 42 °C, hardly any fluorescence was observed and the fusion proteins were detected on immunoblots as a single band that is predicted to be misfolded.

A temperature of 20 °C in combination with the strain BW25113B and an inducer concentration of 0.001% (w/v) L-arabinose was subsequently chosen as default condition to test the functionality of the MP-GFP-ErmC constructs in a 96-well format, that is, by comparing growth on selective media (with varying concentrations of erythromycin). The growth of cells in the presence of erythromycin is an indicator of the activity of ErmC in the fusion protein. In the presence of erythromycin without induction, none of the constructs supported growth in erythromycin-containing medium. With 0.001% (w/v) L-arabinose and in the presence of erythromycin, the cells expressing the target fusions survived. As shown in Fig. 1b, cells overexpressing soluble protein GerAC reached the highest cell density, that is, $A_{600} \sim 0.9$ at 50 mg/ml of erythromycin. Cells expressing GltP, YidC and OxaA reached an intermediate $A_{600}$ of $\sim 0.5$, whereas poorly expressed membrane proteins such as BcaP, LacS and LacClA and PacL supported growth up to $A_{600}$ of $\sim 0.2$. Whole-cell fluorescence of the cultures expressing the corresponding constructs is shown in Fig. 1c. A scatterplot of whole-cell fluorescence versus final $A_{600}$ of the cultures is shown for each construct in Fig. 1d, which indicates a moderate linear dependence in the Pearson product-moment correlation analysis ($r = 0.704$, $p = 0.051$, $n = 8$). We stress that the correlation between expression and $A_{600}$ is not expected to be strong as in general different (membrane) proteins impose different degrees of toxicity to the cell, whereas expression may continue when growth is inhibited. Nevertheless, our selection system effectively discriminates between cells expressing high levels of target membrane protein and those expressing little or no protein.

**Selection of improved expression hosts**

For the selection of BW25113B derivatives with improved expression properties, we used GltP and BcaP as target proteins and thus constructed pBADcGltP–GFP–ErmC and pBADcBcaP–GFP–ErmC. BcaP is from the Gram-positive bacterium L. lactis and GltP is from E. coli, which represent cases of heterologous and homologous expression, respectively. BcaP and GltP belong to different structural classes of membrane proteins [39,40]. Furthermore, BcaP expresses poorly and GltP expresses relatively well in E. coli BW25113B (Fig. 1b). E. coli BW25113B bearing a plasmid for expression of the different MP-GFP-ErmC fusions was grown in LB plus ampicillin and kanamycin, and the concentration of erythromycin was gradually increased. Cells were transferred to fresh medium every 48 h. Cultures resistant to 100 μg/ml erythromycin were plated on LB agar with 100 μg/ml erythromycin to obtain individual clones, which were then screened for folded membrane protein by monitoring the GFP fluorescence in whole cells (data not shown) and analyzing selected samples on SDS-PAGE to confirm full-length expression of MP-GFP-ErmC fusions. Finally, four clones (NG2, NG3, NG5 and NG6) showing increased erythromycin resistance and increased GFP fluorescence were selected for a more complete analysis.

The plasmids in the four strains were isolated and analyzed by DNA sequencing. Coding sequences of the membrane protein fusions and their promoters were not mutated. Their copy numbers were also the same as that of the original plasmids. This suggests that the improvement in membrane protein expression is due to changes in the hosts rather than a genotypic change in the plasmids.

**Protein expression in the evolved E. coli strains**

Next, we cured the evolved strains for their plasmids and retransformed the strains with pBADcBcaP–GFP–ErmC, pBADcGltP–GFP–ErmC and pBADcLacS and LacClA–GFP–ErmC to confirm that the acquired property for enhanced expression is located on the E. coli chromosome and to check for differences in expression patterns. The expression levels and global folding of the proteins were tested by whole-cell fluorescence (total GFP fluorescence), in-gel fluorescence (to discriminate between full-length proteins and possible breakdown products) and anti-His tag immunoblots (to discriminate between folded and misfolded fusion proteins [36]). Figure 2 shows that, in each of the NG strains, the amount of fluorescent (folded) protein and the ratio of folded over misfolded fusion proteins increased up to 10-fold relative to the parental strain BW25113B. Among the evolved strains, NG3 performed best overall, whereas NG2 did better in the overexpression of GltP as compared to BcaP. We note that NG2 and NG3 were selected using GltP as target membrane protein, whereas NG5 and NG6 were obtained with BcaP. Finally, we tested the expression of two soluble reporters, which are GerAC and LmoSBP, and we observed that the level
of GerAC was similarly high in BW25113B and the NG strains, whereas the production of LmoSBP was significantly increased in the evolved *E. coli* strains (data not shown).

**Optimal expression conditions**

To determine the optimal expression conditions, we performed several small-scale expression studies. First, the optimal concentration of inducer, L-arabinose, was determined. Figure 3 shows the expression levels in the evolved and parental strains. NG3 and NG5 show higher levels of expression compared to BW25113B, but also, the sensitivity for L-arabinose is shifted. In the case of NG3 and NG5, 0.01% (w/v) or higher concentrations of L-arabinose were required for maximal expression. We observed that the expression levels increase with induction times. The highest levels of protein were generally found after 16 h of induction, but in most cases, high levels were already obtained after 5–6 h (Fig. 4).

**DNA sequencing of the evolved *E. coli* strains**

Whole-genome sequencing of the parent strain BW25113B and the four evolved NG strains was performed to explore the mechanism responsible for improved membrane protein production. The overall sequence coverage was more than 500-fold across the entire genome. The parent strain BW25113B was found to have 22 differences (mostly point mutations, two insertions) relative to the sequence of *E. coli* K-12 sub-strain MG1655 (GenBank Accession No. U00096.2). When compared to BW25113B, the evolved NG strains carried two or three mutations. Remarkably, each of the NG strains had a mutation associated with *hns*, in the promoter region and/or the coding region. The *hns* mutation in NG5 and NG6 leads to deletion of Thr25, Leu26, Glu27 and Glu28, which reside in the dimerization domain of H-NS (PDB ID: 1NI8) [41]. In NG2 and NG3, codon 126 of *hns*, coding for glutamine, is mutated to a stop codon, which results in a truncation of 12 residues at the C-terminus of the H-NS protein. In addition, NG2 has a mutation in the *hns* promoter.
The mutations found in the four evolved *E. coli* strains are summarized in Table 1 and further discussed below. The striking finding is that, in independent evolution experiments with different target proteins, *hns* is mutated. This prompted us to investigate the role of H-NS in more detail.

**Construction and characterization of *hns* deletion strains**

To investigate the function of *hns* in the increase of membrane protein expression, we deleted the coding region of *hns* from the genome of *E. coli* strains BW25113B, its parent BW25113 and MG1655. *hns* is not an essential gene for *E. coli* [42], and we could successfully generate *hns* null strains. We note that *hns* null strains do grow slower and that the final cell densities are lower when compared to the corresponding parental strains. On the basis of whole-cell fluorescence and gel-based analysis, the *hns* null strains of BW25113 and MG1655 expressed up to 2-fold more correctly folded BcaP (Fig. 5), GltP and LacSΔIIA (Fig. S3) than the corresponding parental strains. However, the expression was never as high as in NG3, NG5 or NG6, suggesting that deletion of residues in the dimerization domain (NG5 and NG6) or truncation of the DNA binding domain (NG2 and NG3) leads to a partially functional H-NS molecule, which is important for high-level expression. We also discuss below the role of the secondary mutations in the NG strains.

**Discussion**

The production of integral membrane proteins in a fully functional state is often a hit-or-miss affair. Ideally, one would like to tackle the expression problem without having to rely on optimization without trial and error. In this paper, we describe a simple and effective strategy to optimize the level of functional membrane protein in *E. coli*. Our approach is based on a double translational fusion reporting system, using the *gfp* and *ermC* genes in tandem and fused 3′ of the target region.
gene. The GFP protein is used to monitor the folding state of the target protein; the ErmC protein (23S ribosomal RNA adenine N-6 methyltransferase) is used to screen for increased functional expression by evolving strains toward increasing concentrations of erythromycin. A manageable screen for expression optimization can be performed in as little as 200 μl of culture and performed for numerous target proteins in parallel. We previously focused on *L. lactis* for enhanced membrane protein expression. The evolved *L. lactis* strains acquired mutations in *nisK*, which increased the transcription from the plasmid-based nisin promoter [37]. We now adopted a similar strategy to evolve *E. coli* for enhanced production of integral membrane proteins, but the method is applicable for any type of protein. When the technology is used in its full potential, we foresee a panel of expression hosts, each evolved and optimized for a given class of proteins or tuned toward a specific protein.

![Fig. 4](image)

**Fig. 4.** Optimal induction time at 0.01% (w/v) L-arabinose. (a) GFP fluorescence of cells carrying pBADcGltP-GFP-ErmC. Samples were taken after 4 h (black) and 16 h (gray) of induction at 25 °C. (b) Expression of GltP in BW25113B, NG2 and NG3 at 0.01% (w/v) L-arabinose and different lengths of induction. In-gel GFP fluorescence and immunoblots probed with an anti-His tag antibody of the same gels are shown. Black and white arrows indicate the positions of non-fluorescent (misfolded) and fluorescent (folded) protein species, respectively. The ratio of folded to misfolded protein is indicated below the immunoblot.
of expression level and the ratio of well-folded over misfolded protein. We show that the fluorescence intensity of the target protein-GFP-ErmC fusion is a good indicator of the level of functional expression of the target protein. Importantly, the whole-cell fluorescence correlates with cell growth in the presence of erythromycin, which offers a simple readout for the selection of better expressing strains.

We obtained several evolved strains of which four were characterized in more detail, that is, NG2, NG3, NG5 and NG6. NG3 proved most effective for all target proteins tested. In general, the evolved strains performed best for the target protein that was used in the initial screening. For instance, in NG2, the increase in expression of GFP was higher than for BcaP, whereas NG5 and NG6 performed relatively best with BcaP. In the past decade, BL21(DE3) derivatives C41(DE3) and C43(DE3), also known as Walker strains [45], have been widely used for the expression of membrane proteins. The strains yield more membrane protein because the acquired mutations lower the efficiency of the T7 transcription machinery. The reduction in synthesis of target protein prevents saturation of the Sec translocon and hence minimizes the toxic effect of overexpression of membrane proteins. As a result, more biomass is produced and more protein is obtained per liter of cell culture [26]. In contrast to the Walker strains, the NG strains show increased expression per unit of biomass. Genome sequencing revealed that the NG strains have different mutations, but remarkably, all have a mutation in hns. H-NS is an abundant nucleoid-associated protein, involved in genome organization and transcriptional silencing of a variety of cellular functions [46–48]. H-NS consists of N- and C-terminal domains separated by a flexible linker region. The N-terminal part is composed of three α helices connected by linkers, and this domain is required for dimerization of the protein [41]. The C-terminal domain of H-NS binds to DNA and preferably to sites rich in AT [47]; the ICGATAAATT sequence has been identified as consensus motif for high-affinity binding [49,50]. H-NS represses transcription by binding to high-affinity sites and cooperative spreading along DNA, thereby occupying the promoter region. As a consequence, it also forms looped structures of DNA [51,52]. In fact, H-NS and other nucleoid-associated proteins (Fis, Dps and StpA) are thought to form large topological domain barriers, thereby determining the chromosome architecture. The interaction of H-NS with DNA is affected by several environmental factors [53], including pH [54], osmolarity [55,56], temperature [57] and growth phase [58], but it differs from gene to gene and, as such, the regulation of gene expression is complex. H-NS is unlikely to have a direct effect on the transcription induction machinery used here to drive the expression of membrane proteins. However, it may affect the topology of the plasmid DNA and indirectly affect transcription. Furthermore, as a global regulator, H-NS may change the expression levels of downstream components in membrane protein biogenesis, such as the Sec translocase and chaperones, which in turn could enhance membrane protein production.

H-NS is also known to affect the expression of drug exporters and contributes to multidrug resistance in E. coli [59]. In our directed evolution system, erythromycin was used to provide selection pressure. The clones isolated may thus have been selected not only on the basis of increased levels of properly folded MP-GFP-ErmC fusions but also through increased erythromycin export capacity. Importantly, in all tests to characterize the evolved strains, erythromycin was no longer present in the medium and the increase in membrane protein expressions was not dependent on the antibiotic.

NG2 and NG3 have a nonsense mutation in hns, which results in the deletion of the C-terminal 12 amino acids from H-NS and most likely a defective transcription factor. NG2 has an additional mutation

### Table 1. Mutations in evolved E. coli strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Mutated gene</th>
<th>Mutation</th>
<th>Amino acid change</th>
<th>Function of protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>NG2</td>
<td>hns</td>
<td>376 C → T</td>
<td>Gln126Stop</td>
<td>Chromosome organization and transcriptional silencing</td>
</tr>
<tr>
<td></td>
<td>hns</td>
<td>−29 G → A</td>
<td>(N/A)</td>
<td>Chromosome organization and transcriptional silencing</td>
</tr>
<tr>
<td></td>
<td>ung</td>
<td>220 G → A</td>
<td>Gly74Ser</td>
<td>Uracil DNA glycosylase that modulates the mutation frequency; controlled by the cpx regulon</td>
</tr>
<tr>
<td></td>
<td>cpxA</td>
<td>524 G → A</td>
<td>Ser175Asn</td>
<td>Histidine sensor kinase involved in cell envelope stress response</td>
</tr>
<tr>
<td>NG3</td>
<td>hns</td>
<td>376 C → T</td>
<td>Gln126 to stop codon</td>
<td>Chromosome organization and transcriptional silencing</td>
</tr>
<tr>
<td></td>
<td>dcm</td>
<td>1337 T → A</td>
<td>Val446Glu</td>
<td>DNA-cytosine methyltransferase</td>
</tr>
<tr>
<td></td>
<td>hns</td>
<td>Deletion from A73 to A84</td>
<td>Unknown</td>
<td>Chromosome organization and transcriptional silencing</td>
</tr>
<tr>
<td>NG5</td>
<td>ydcU</td>
<td>348 C → A</td>
<td>(Silent)</td>
<td>Putative ABC transporter for spermidine and putrescine</td>
</tr>
<tr>
<td></td>
<td>yffN</td>
<td>195 C → T</td>
<td>(Silent)</td>
<td>Unknown</td>
</tr>
<tr>
<td>NG6</td>
<td>hns</td>
<td>Deletion from A73 to A84</td>
<td>Deletion of Thr25, Leu26, Glu27 and Glu28</td>
<td>Chromosome organization and transcriptional silencing</td>
</tr>
<tr>
<td>ydcU</td>
<td>348 C → A</td>
<td>(Silent)</td>
<td>(Silent)</td>
<td>Putative ABC transporter for spermidine and putrescine</td>
</tr>
</tbody>
</table>
at position −29 of the coding region, that is, the promoter region of hns. This mutation may change the expression of mutated H-NS proteins in NG2 and hence contribute to the difference in expression performance between NG2 and NG3. NG5 and NG6 have a deletion of 12 nucleotides that results in a deletion of 4 residues in the N-terminal domain of H-NS and most likely affects the dimerization of the protein. As both mutations may result in either total or partial loss of H-NS function, we hence constructed strains with the coding sequence of hns completely replaced by a drug resistance marker gene. Indeed, the hns null strains showed a 1.5- to 2-fold increase in membrane protein biogenesis but the extent of amplification was significantly less than that in the evolved strains.

Fig. 5. Expression of BcaP-GFP-ErmC in different hns deletion strains and the evolved strains. We used 0.01% (w/v) L-arabinose for induction at $A_{600}$ of 0.5–0.6, and the expression was continued at 25 °C for 16 h. (a) Whole-cell fluorescence. Values are normalized on the basis of $A_{600}$. The error bars show standard error of the mean from three separate experiments. *$P = 0.119$, **$P = 0.024$ and ***$P = 0.002$ as determined by the paired T-test. (b) In-gel GFP fluorescence and anti-His immunoblot of the same gel from different strains expressing BcaP-GFP-ErmC; to improve the illustration, we rearranged the lanes but all originate from the same immunoblot and the intensities can be compared quantitatively. The culture and expression conditions are the same as in (a). Black and white arrows indicate the positions of non-fluorescent (misfolded) and fluorescent (folded) protein species, respectively. The ratio of folded to misfolded protein is indicated below the immunoblot. Representative data from the three separate experiments are shown.
What about the secondary mutations in the NG strains? NG2 has substitution mutations in ung and cpxA in addition to a nonsense mutation in hns. CpxA is the histidine sensor kinase of a two-component signal transduction system. The Cpx two-component regulatory systems respond to pH and osmotic stress, as well as aggregated and misfolded proteins [60]. Cpx activates the expression of protein folding catalysts and protein degradation factors [61] and could play a role in membrane protein expression. Phosphorylated CpxR, the response regulator of the Cpx system, negatively regulates ung transcription [62]. ung codes for uracil DNA glycosylase, which removes uracil from DNA via base excision repair and thus diminishes the mutation level of E. coli. We consider it unlikely that the mutation in ung has contributed to the phenotype of NG2 and attribute the properties of this strain primarily to altered DNA binding properties of truncated H-NS.

NG3 has an additional mutation in dcm, which codes for DNA-cytosine methyltransferase, an enzyme involved in DNA methylation. Methylated cytosine is subject to deamination, which causes C-to-T mutations. This mismatch is repaired by the very short patch (VSP) mismatch repair pathway. Vsr is an essential component of VSP, and its activity counterbalances the mutagenesis associated with dcm activity [63]. The chromosomal arrangement of dcm and vsr is unusual, and the 5’ end of vsr overlaps the 3’ end of dcm in the +1 reading frame [64]. The two genes are probably transcribed as a single mRNA, with translation of vsr depending on translation of dcm [65]. This gene arrangement may assure that Vsr is produced along with Dcm and may minimize the mutagenic effects of cytosine methylation. NG3 has a mutation in the 3’ region of dcm, which might affect VSP-mediated mismatch repair. Since NG3 shows the highest levels of expression of different membrane proteins and its phenotype is stronger than that of NG2 or the isogenic deletion strain, we feel that the dcm mutation may act synergistic to the hns mutation.

NG5 and NG6 have the same mutation in ydcU in addition to a mutation that affects the dimerization of H-NS. YdcU is the membrane component of a binding-protein-dependent ABC transport system that takes up spermidine/putrescine. However, the mutation in ydcU is silent, and most likely, the mutation has not contributed to the phenotype of NG5 and NG6. Since NG5 and NG6 have the same mutations in hns and ydcU, it seems likely that the two strains have not evolved independently.

In conclusion, the strategy presented here allows rapid screening of amplified expression of almost any (membrane) protein, which will generate an even larger variety of better expressing hosts. Combined with the ease nowadays to sequence prokaryotic genomes, it will generate a wealth of information on what limits the functional expression of membrane proteins. In this study, we already identified the nucleoid-associated protein H-NS as an important factor in the functional overexpression of membrane proteins. The further exploration of the selection method will eventually lead to rules regarding expression bottlenecks for different classes of integral membrane proteins. The strains isolated and characterized will be made available for use by the scientific community.

### Materials and Methods

#### Strains and cultivation conditions for protein expression

The strains used in this study are listed in Table S1. For cloning procedures, *E. coli* MC1061 [66] was used as a host, and *E. coli* BW25113 ΔacrB ΔemrE ΔmdfA∷kan [38] (BW25113B) was used as host for the selection of strains with improved (evolved) expression performance. *E. coli* BW25113B is extremely sensitive to erythromycin because the genes encoding three major drug extrusion systems have been inactivated (ΔacrB ΔemrE ΔmdfA∷kan). *E. coli* strains carrying pBADcLIC derivatives were cultivated under vigorous aeration at 37 °C in Luria broth supplemented with 100 μg/ml ampicillin. In the case of *E. coli* BW25113B and derivatives, cells were cultured in Luria broth supplemented with 100 μg/ml ampicillin plus 5 μg/ml kanamycin at 37 °C (unless indicated) in shake flasks rotated at 200 rpm.

#### Construction of GFP-ErmC fusion proteins

DNA manipulations were performed according to standard protocols. The vector pBADcLIC-ErmC was obtained by insertion of the cLIC-ErmC cassette from pREcLIC-ErmC [37] into pBADcLIC [67] as NcoI-HindIII fragment. Plasmid pBADcLIC–GFP–ErmC was constructed in two steps. First, the GFP gene was amplified from pBADcLIC–GFP, using primer-introduced sites at the 5′ (SwaI) and 3′ (EcoRI-XbaI) ends; the gene was subsequently cloned as a SwaI-XbaI fragment in the pBADcLIC vector, yielding pBADcLIC–GFP+EcoRI (the EcoRI site being created by the reverse primer). A EcoRI-XbaI fragment, corresponding to the ermC-His sequence, amplified from pREcLIC–ErmC, was cloned into pBADcLIC–GFP+EcoRI, yielding pBADcLIC–GFP–ErmC.

#### Target proteins

The target genes coding for BcaP (branched-chain amino acid permease from *L. lactis*), LacSα1α (a C-terminal truncation mutant of the lactose transporter from *S. thermophilus*), PcaL (putative P-type cation transporter from *L. lactis*), YidC (preprotein translocase from *L. lactis*), GerAC (spore germination factor from *Bacillus subtilis*), OxaA (YidC-like protein from *L. lactis*), GltP (glutamate transporter from *E. coli*) and Sav1866
(multidrug ABC transporter from *Staphylococcus aureus*) were ligated into pBADcLIC–GFP–ErmC, using the ligation-independent cloning method [67]. In all cases, the target protein was expressed with a C-terminal tobacco etch virus protease cleavage site (sequence ENLYFQG) followed by the GFP, ErmC and a 10× His tag.

**Selection and screening**

*E. coli* BW25113B transformed with pBADcGltP–GFP–ErmC, pBADcBcaP–GFP–ErmC or pBADcLacSΔIIA–GFP–ErmC were grown in LB supplemented with 100 μg/ml ampicillin plus 5 μg/ml kanamycin. The temperature was lowered from 37 °C to 25 °C after the *A*₆₀₀ had reached 0.5–0.6 (all absorbance values were measured through a 1-cm light path), and after temperature equilibration, protein expression was induced by the addition of 0.01% (w/v) of L-arabinose. After 2 h of induction, a 2% (v/v) inoculum was introduced into fresh LB supplemented with 100 μg/ml ampicillin and 5 μg/ml kanamycin plus 5 μg/ml erythromycin.

Cultivation was continued for 48 h, after which the cells were diluted into fresh medium with an elevated erythromycin concentration [always using a 2% (v/v) culture inoculum]. The erythromycin concentration was gradually increased from 5 to 10, to 20, to 50, to 100, to 200 μg/ml. After that, cells were plated onto LB agar supplemented with 100 μg/ml ampicillin, 5 μg/ml kanamycin and 100 μg/ml erythromycin plus 0.01% (v/v) L-arabinose. Isolated colonies were selected for further analysis.

**Whole-cell fluorescence measurements**

For whole-cell fluorescence measurements, cells expressing membrane protein-GFP-ErmC fusions were harvested, washed and subsequently resuspended to *A*₆₀₀ = 3 (1 mg/ml of total protein) in 100 mM KPi (pH 7). GFP emission was measured using a BioTek FL600 microplate fluorescence reader with an excitation filter of 485/20 nm and an emission filter of 530/25 nm. For direct comparison of the expression levels, the GFP fluorescence data were normalized to the same *A*₆₀₀. Background fluorescence levels were assessed by measuring whole-cell fluorescence of uninduced cells, and these values were subtracted.

**In-gel fluorescence and immunoblotting**

Whole-cell samples corresponding to 3 *A*₆₀₀ units (1 mg of total protein) were resuspended in 400 μl ice-cold 50 mM KPi (pH 7.2), 1 mM MgSO₄, 10% glycerol and 1 mM PMSF plus trace amounts of DNase. Glass beads (300 mg, 0.1 mm diameter) were added, and samples were shaken in a Tissue Lyser LT (Qiagen) at 50 oscillations per second for 5 min. Aliquots (20 μl, 50 μg protein) were taken, and 5 μl of 5× protein sample buffer (120 mM Tris–HCl (pH 6.8), 50% (v/v) glycerol, 100 mM DTT and 2% (w/v) SDS plus 0.1% (w/v) bromophenol blue) was added (2 mg/ml final). Samples were stored on ice until use. Protein samples [10 μl (20 μg) for GIP and LacSΔIIA; 15 μl (30 μg) for BcaP] were analyzed by 10% SDS-PAGE electrophoresis and in-gel GFP fluorescence, using a Fujifilm LAS-3000 imaging system and AIDA software (Raytest; Isotopenmessgeräte, GmbH). Subsequently, the proteins in the gels were transferred onto polyvinylidene difluoride membranes and probed with anti-hexa-His tag primary antibody (Amersham Pharmacia Biotech) and alkaline phosphatase conjugated anti-mouse IgG secondary antibody (Sigma). Chemiluminescence detection was performed using the Western-Light kit (Tropix, Inc.) and quantified using the Fujifilm LAS-3000 imaging system.

**Erythromycin resistance**

The level of functional ErmC protein was estimated by quantification of bacterial survival as a function of erythromycin concentration. *E. coli* BW25113B cells expressing the different membrane protein-GFP-ErmC fusion were inoculated [2% (v/v)] into a fresh LB liquid medium with erythromycin concentrations varying from 0 to 100 μg/ml and were subsequently incubated overnight. The final optical density (*A*₆₀₀) reached by the culture was the parameter utilized to determine ErmC activity.

**Plasmid copy number and plasmid curing**

The plasmid copy number was determined by isolation of plasmid DNA, using the Promega miniprep kit and estimation of the plasmid DNA concentration by Nanodrop readout at 260 nm. For curing of evolved strains of pBADc-MP-GFP-ErmC, the cells were grown at 42 °C in LB supplemented with 5 μg/ml kanamycin and subsequent transfer [0.1% (v/v) inoculum] into pre-warmed media. After 25 rounds of sub-culturing, the cells were plated on LB agar plates and isolated colonies were further characterized by culturing on LB agar plates supplemented with and without 100 μg/ml ampicillin. Single colonies of cured clones were purified further, and the cells were retransformed with the original plasmids and screened for expression performance.

**Genome sequencing and data analysis**

Genomic DNA was purified using the Qiagen Genomic DNA Purification Kit. Shotgun DNA libraries were generated according to the manufacturer’s sample preparation protocol for genomic DNA. Briefly, 1–5 μg of genomic DNA was randomly sheared using nebulizers and the ends were repaired using polynucleotide kinase and Klenow enzyme. The 5′ ends of the DNA fragments were phosphorylated and a single adenine base was added to the 3′ ends using Klenow exonuclease. Following ligation of a pair of Illumina adaptors to the repaired ends, we amplified the DNA in 10 cycles, using adaptor primers (Illumina, San Diego, CA), and we isolated fragments around 250 bp long from agarose gels. Sequencing libraries were quantified with a Bio-Rad Experion Analyzer and the Picogreen fluorescence assay. Cluster generations were performed on an Illumina cluster station using 4 pmol of sequencing libraries. Seventy-six cycles of sequencing were carried out using the Illumina GenomeAnalyzerII system according to the manufacturer’s specifications. Sequencing analysis was first performed using the Illumina analysis pipeline. The output of the Illumina analysis pipeline was fed into CLC Bio Software (Aarhus, Denmark). At the assembly stage, sequence reads
of each mutant were aligned to the previously assembled wild-type (reference) genome.

**Construction of hns deletion strains**

The coding region of hns was removed as described by Datsenko and Wanner [68] with minor modifications. In brief, genome modification of E. coli was performed by replacing a locus by a linear DNA fragment. It contained a selection marker gene, flanked by 50-bp sequences homologous to the upstream and downstream region of hns and Fip recognition targets at both ends. For knocking out hns in BW25113 and MG1655, we generated the DNA fragment by PCR using pEAW507 (gift from Andrew Robinson, University of Groningen) as template, resulting in the recombination of the double-stranded DNA fragment with the kanamycin resistance cassette. For kanamycin-resistant BW25113B, the kanamycin resistance cassette of pEAW507 was first replaced by the chloramphenicol resistance cassette from pH29 (Andrew Robinson). The recombining DNA was generated by PCR using the same pair of primers. E. coli strains BW25113, BW25113B and MG1655 were transformed by pKD46, an expression plasmid with P_{BAD} driven phage lambda RED recombinase genes. The transformed cells were then grown in SOB at 30 °C and induced by 0.2% antibiotic-based selection; evolved strains; functional overexpression

† N.G., D.M.L. and F.Y.H. contributed equally to this work.

**Abbreviations used:**

GFP, green fluorescent protein; VSP, very short patch.

**References**


Evolved Escherichia coli for protein expression


